On Local Symmetries and Universality in Cellular Automata

Laurent Boyer, Guillaume Theyssier

Symposium on Theoretical Aspects of Computer Science 2009 (Freiburg), pp. 195–206
www.stacs-conf.org

ON LOCAL SYMMETRIES AND UNIVERSALITY IN CELLULAR AUTOMATA

LAURENT BOYER [1] AND GUILLAUME THEYSSIER [1]

[1] LAMA (CNRS, Université de Savoie),
Campus Scientifique, 73376 Le Bourget-du-lac cedex FRANCE
URL: `http://www.lama.univ-savoie.fr`
E-mail address, L. Boyer: `laurent.boyer@univ-savoie.fr`

E-mail address, G. Theyssier: `guillaume.theyssier@univ-savoie.fr`

Abstract. Cellular automata (CA) are dynamical systems defined by a finite local rule but they are studied for their global dynamics. They can exhibit a wide range of complex behaviours and a celebrated result is the existence of (intrinsically) universal CA, that is CA able to fully simulate any other CA. In this paper, we show that the asymptotic density of universal cellular automata is 1 in several families of CA defined by local symmetries. We extend results previously established for captive cellular automata in two significant ways. First, our results apply to well-known families of CA (e.g. the family of outer-totalistic CA containing the Game of Life) and, second, we obtain such density results with both increasing number of states and increasing neighbourhood. Moreover, thanks to universality-preserving encodings, we show that the universality problem remains undecidable in some of those families.

Introduction and definitions

The model of cellular automata (CA) is often chosen as a theoretical framework to study questions raised by the field of complex systems. Indeed, despite their formal simplicity, they exhibit a wide range of complexity attributes, from deterministic chaos behaviours (*e.g.* [3]) to undecidability in their very first dynamical properties (*e.g.* [2]). One of their most important feature is the existence of universal CA. Universality in CA is sometimes defined by an adaptation from the model of Turing machines and sequential calculus. But a stronger notion, intrinsic to the model of CA, has emerged in the literature [7]: a CA is *intrinsically universal* if it is able to fully simulate the behaviour of any other CA (even on infinite configurations).

Besides, when it comes to modelling [1] or experimental studies [10, 11], most works focus on some particular syntactical families (elementary CA, totalistic CA, etc), either to reduce the size of the rule space to explore, or to match hypothesis of the studied phenomenon at microscopic level (*e.g.* isotropy).

1998 ACM Subject Classification: F.1.1, G.2.1, F.4.3.
Key words and phrases: cellular automata, universality, asymptotic density.

In a word, CA are known for their general ability to produce complex global behaviours, but local rule considered in practice are often very constrained. This paper studies the link between syntactical restriction on CA local rules and typical global behaviours obtained. It establishes a probabilistic result: for various symmetry criterions over local rules, randomly choosing a local rule within the symmetric ones yields almost surely universal CA. Meanwhile, the universality problem is shown to remain undecidable even restricted to symmetric rules (for some of the symmetry criterions).

A family of CA defined by a simple syntactical constraint (namely *captive* CA) and containing almost only universal CA has already been proposed by one of the authors [9], but the present paper goes further. First, it generalises the probabilistic framework: the neighbourhood of CA is no longer fixed as it was needed in [9]. Second, it considers well-known families of CA (*e.g.* totalistic or outer-totalistic CA) and generalisations of them, namely *multiset* CA, which are meaningful for modelling (they are 'isotropic' CA).

After having recalled standard definitions about CA (end of this section), section 1 presents the families considered in this paper. Then, section 2 defines intrinsic universality and the simulation relation involved in that notion. Section 3 gives the probabilistic setting of the paper and establishes the main probabilistic results. Finally, section 4 is dedicated to existence proofs of universal CA in the families considered. Combined with probabilistic results, it proves that almost all CA are universal in those families.

Definitions and notations. In this paper, we adopt the setting of one-dimensional cellular automata. Formally, a CA is a 3-uple $\mathcal{A} = (n, k, \delta_\mathcal{A})$ where n and k are positive integers, respectively the size of the state set $\mathcal{Q}_n = \{1, \ldots, n\}$ and of the neighbourhood $[\![-\lfloor\frac{k-1}{2}\rfloor; \lfloor\frac{k}{2}\rfloor]\!]$, $\delta_\mathcal{A} : \mathcal{Q}_n^k \to \mathcal{Q}_n$ is the *local transition function.*

A coloring of the lattice $\mathbb{Z}$ with states from $\mathcal{Q}_n$ (*i.e.* an element of $\mathcal{Q}_n^\mathbb{Z}$) is called a *configuration.* To $\mathcal{A}$ we associate a global function $G_\mathcal{A}$ acting on configurations by synchronous and uniform application of the local transition function. Formally, $G_\mathcal{A} : \mathcal{Q}_n^\mathbb{Z} \to \mathcal{Q}_n^\mathbb{Z}$ is defined by: $G_\mathcal{A}(x)_z = \delta_\mathcal{A}(x_{z-\lfloor\frac{k-1}{2}\rfloor}, \ldots, x_{z+\lfloor\frac{k}{2}\rfloor})$ for all $x \in \mathcal{Q}_n^\mathbb{Z}$ and $z \in \mathbb{Z}$.

The local function $\delta_\mathcal{A}$ naturally extends to $\mathcal{Q}_n^*$, the set of finite words over alphabet $\mathcal{Q}_n$ (with $\delta_\mathcal{A}(u)$ being the empty word if $|u| < k$). For $p \in \mathbb{N}$, this function maps an element of $\mathcal{Q}_n^{p+k}$ to an element of $\mathcal{Q}_n^{p+1}$.

The size of $\mathcal{A} = (n, k, \delta_\mathcal{A})$ is the pair (n, k). The set of all CA is denoted by $\mathbf{CA}$, and the set of all CA of size (n, k) by $\mathbf{CA}_{n,k}$. Moreover for any set $\mathcal{F} \subseteq \mathbf{CA}$, $\mathcal{F}_{n,k}$ is defined by $\mathcal{F}_{n,k} = \mathcal{F} \cap \mathbf{CA}_{n,k}$. Formally a CA is a 3-uple but, to simplify notation, we sometimes consider that $\mathcal{F}_{n,k}$ is a set of local functions of type $\mathcal{Q}_n^k \to \mathcal{Q}_n$.

This paper will intensively use (finite) multisets. A multiset M of elements from a set E is denoted by $M = \{\!\{(e_1, n_1), \ldots, (e_p, n_p)\}\!\}$ where a pair $(e_i, n_i) \in E \times \mathbb{N}$ denotes an element and its multiplicity. The cardinality of M is $|M| = \sum_i n_i$. The cardinality notation is the same for sets.

1. Families of CA with Local Symmetries

In this section, we define various families of CA characterised by some local symmetry. 'Symmetry' must be taken in a broad sense since it may concern various aspects of the local function. We first consider families where the local function does not depend on the exact configuration of the neighbourhood (a k-uple of states) but only on a limited amount of information extracted from this configuration.

MultiSet CA. Multiset cellular automata are cellular automata with a local rule invariant by permutation of neighbours. Equivalently, they are CA whose local function depends only on the multiset of states present in the neighbourhood. Formally, $\mathcal{A} \in \mathbf{CA}_{n,k}$ is *multiset*, denoted by $\mathcal{A} \in \mathbf{MS}_{n,k}$, if for any permutation π of $\{1 \ldots k\}$, the local function $\delta_{\mathcal{A}}$ satisfies

$$\forall a_1, \ldots, a_k \in \mathcal{Q}_n : \delta_{\mathcal{A}}(a_1, \ldots, a_k) = \delta_{\mathcal{A}}(a_{\pi(1)}, \ldots, a_{\pi(k)}).$$

Set CA. Set CA are a special case of multiset CA: they are CA whose local function depends only on the set of states present in the neighbourhood. Formally, $\mathcal{A} \in \mathbf{CA}_{n,k}$ with arity k is a *set CA*, denoted by $\mathcal{A} \in \mathbf{Set}_{n,k}$, if

$$\forall u, v \in \mathcal{Q}_n^k : \{u_1, \ldots, u_k\} = \{v_1, \ldots, v_k\} \Rightarrow \delta_{\mathcal{A}}(u) = \delta_{\mathcal{A}}(v).$$

Note that for fixed n, there is a constant N such that, for all k, $|\mathbf{Set}_{n,k}| \leq N$. Thus there is no hope that the asymptotic density of a non-trivial property for fixed n be 1 for family **Set**.

Totalistic CA. Totalistic CA are also a special case of Multiset CA: they are CA whose local functions depends only on the sum of the neighbouring states. Formally, $\mathcal{A} \in \mathbf{CA}_{n,k}$ k is *totalistic*, denoted by $\mathcal{A} \in \mathbf{Tot}_{n,k}$, if

$$\forall u, v \in \mathcal{Q}_n^k : \sum_{i=1}^{k} u_i = \sum_{i=1}^{k} v_i \Rightarrow \delta_{\mathcal{A}}(u) = \delta_{\mathcal{A}}(v).$$

Partial Symmetries. We can consider weaker forms of each family above, by excluding some neighbours from the 'symmetry' constraint and treating them as a full dependency in the local function. For instance, we define the set of *outer-multiset* CA as those with a local rule depending arbitrarily on a small central part of their neighbourhood and on the multiset of other neighbouring states. Formally, for any k', $0 \leq k' \leq k$, $\mathbf{O}_{k'}\mathbf{MS}_{n,k}$ is the set of CA with n states, arity k and such that for any permutation π of $\{1 \ldots k - k'\}$ and any $a_1, \ldots, a_{k-k'}, b_1, \ldots, b_{k'} \in \mathcal{Q}_n$ we have:

$$\begin{aligned}\delta_{\mathcal{A}}(a_1, \ldots, a_{\lfloor (k-k')/2 \rfloor}, &b_1, \ldots, b_{k'}, a_{\lfloor (k-k')/2 \rfloor + 1}, \ldots, a_{k-k'}) \\ &= \delta_{\mathcal{A}}\big(a_{\pi(1)}, \ldots, a_{\pi(\lfloor (k-k')/2 \rfloor)}, b_1, \ldots, b_{k'}, a_{\pi(\lfloor (k-k')/2 \rfloor + 1)}, \ldots, a_{\pi(k-k')}\big).\end{aligned}$$

We define in a similar way *outer-totalistic* and *outer-set*, and denote them by $\mathbf{O}_{k'}\mathbf{Tot}_{n,k}$ and $\mathbf{O}_{k'}\mathbf{Set}_{n,k}$ respectively. Note that what is classically called *outer-totalistic* is exactly the family $\mathbf{O}_1\mathbf{Tot}_{n,k}$.

State symmetric CA. Families above are variations around the invariance by permutations of neighbours. State symmetric CA are CA with a local function invariant by permutation of the state set. Formally, a CA $\mathcal{A} \in \mathbf{CA}_{n,k}$ is *state symmetric*, denoted by $\mathcal{A} \in \mathbf{SS}_{n,k}$, if for any permutation π of $\mathcal{Q}_n$ we have:

$$\forall a_1, \ldots, a_k : \delta_{\mathcal{A}}(a_1, \ldots, a_k) = \pi^{-1}\big(\delta_{\mathcal{A}}(\pi(a_1), \ldots, \pi(a_k))\big).$$

Note that we have a situation similar to the case of **Set**: for fixed k, there is a constant K such that, for all n, $|\mathbf{SS}_{n,k}| \leq K$. Thus their is no hope that the asymptotic density of a non-trivial property for fixed k be 1 in state symmetric CA.

Captive CA. Finally, we consider the family of captive CA already introduced in [8]: they are CA where the local function is constrained to produce only states already present in the neighbourhood. Formally, a CA $\mathcal{A} \in \mathbf{CA}_{n,k}$ is *captive*, denoted by $\mathcal{A} \in \mathbf{K}_{n,k}$, if:

$$\forall a_1, \ldots, a_k : \delta_{\mathcal{A}}(a_1, \ldots, a_k) \in \{a_1, \ldots, a_k\}.$$

The following lemma shows a strong relationship between captive and state symmetric CA.

Lemma 1.1. *Let n and k be such that $1 \leq k \leq n-2$. Then we have $\boldsymbol{SS}_{n,k} \subseteq \boldsymbol{K}_{n,k}$.*

Combining symmetries. In the sequel, we will often consider intersections of two of the families above. Note that all intersections are generally non-trivial. However, for the case of $\mathbf{Tot}_{n,k}$ and $\mathbf{K}_{n,k}$, the intersection is empty as soon as there exists two k-uple of states with disjoint support but with the same sum, because the 'captive' constraint forces the two corresponding transitions to be different whereas the 'totalistic' constraint forces them to be equal. This happens for instance when $n \geq 3$ and k is even with k-uples $(1, 3, 1, 3, \ldots, 1, 3)$ and $(2, 2, 2, \ldots, 2)$.

2. Simulations and Universality

The property we are mostly interested in is intrinsic universality (see [7] for a survey on universality). To formalize it, we first define a notion of simulation.

A CA $\mathcal{A}$ is a *sub-automaton* of a CA $\mathcal{B}$, denoted $\mathcal{A} \sqsubseteq \mathcal{B}$, if there is an injective map φ from A to B such that $\overline{\varphi} \circ G_{\mathcal{A}} = G_{\mathcal{B}} \circ \overline{\varphi}$, where $\overline{\varphi} : A^{\mathbb{Z}} \to B^{\mathbb{Z}}$ denotes the uniform extension of φ to configurations. We sometimes write $\mathcal{A} \sqsubseteq_{\varphi} \mathcal{B}$ to make φ explicit. This definition is standard but yields to a very limited notion of simulation: a given CA can only admit a finite set of (non-isomorphic) CA as sub-automata. Therefore, following works of J. Mazoyer and I. Rapaport [4] and later N. Ollinger [5, 7], we will consider the following notion of simulation: a CA $\mathcal{A}$ simulates an AC $\mathcal{B}$ if some *rescaling* of $\mathcal{A}$ is a sub-automaton of some *rescaling* of $\mathcal{B}$. The ingredients of the rescalings are simple: packing cells into blocs, iterating the rule and composing with a translation (formally, we use shift CA σ_z, $z \in \mathbb{Z}$, whose global rule is given by $\sigma(c)_x = c_{x-z}$ for all $x \in \mathbb{Z}$). Formally, given any state set Q and any $m \geq 1$, we define the bijective packing map $b_m : Q^{\mathbb{Z}} \to \left(Q^m\right)^{\mathbb{Z}}$ by:

$$\forall z \in \mathbb{Z} : \big(b_m(c)\big)(z) = \big(c(mz), \ldots, c(mz + m - 1)\big)$$

for all $c \in Q^{\mathbb{Z}}$. The rescaling $\mathcal{A}^{<m,t,z>}$ of $\mathcal{A}$ by parameters m (packing), $t \geq 1$ (iterating) and $z \in \mathbb{Z}$ (shifting) is the CA of state set Q^m and global rule:

$$b_m \circ \sigma_z \circ G_{\mathcal{A}}^t \circ b_m^{-1}.$$

With these definitions, we say that $\mathcal{A}$ simulates $\mathcal{B}$, denoted $\mathcal{B} \preccurlyeq \mathcal{A}$, if there are rescaling parameters m_1, m_2, t_1, t_2, z_1 and z_2 such that $\mathcal{B}^{<m_1,t_1,z_1>} \sqsubseteq \mathcal{A}^{<m_2,t_2,z_2>}$. In the sequel, we will discuss *supports* of simulations, *i.e.* sets of configurations on which simulations occur. If $\mathcal{B}^{<m_1,t_1,z_1>} \sqsubseteq_{\varphi} \mathcal{A}^{<m_2,t_2,z_2>}$, the support of the simulation is the set of configuration of $\mathcal{A}$ defined by $b_{m_2}^{-1} \circ \overline{\varphi} \circ b_{m_1}(\mathcal{Q}_{\mathcal{B}}^{\mathbb{Z}})$. It is a subshift: a closed shift-invariant set of configurations. In the sequel we denote by $\mathcal{B} \preccurlyeq_X \mathcal{A}$ the fact that $\mathcal{A}$ simulates $\mathcal{B}$ on support X.

Once formalised the notion of simulation, we naturally get a notion of universality: CA able to simulate any other CA, denoted $\mathcal{A} \in \mathcal{U}$. This notion associated to $\preccurlyeq$ is called *intrinsic universality* in the literature (see [7]). Actually, an intrinsically universal CA $\mathcal{A}$

has the following stronger property (see [7, 5]): for all $\mathcal{B}$, there are rescaling parameters m, t and z such that $\mathcal{B} \sqsubseteq \mathcal{A}^{<m,t,z>}$.

3. Asymptotic Density and Monotone Properties

3.1. Asymptotic density

When considering a property $\mathcal{P}$ and a family $\mathcal{F}$ (two sets of CA), we can define the probability of $\mathcal{P}$ in $\mathcal{F}_{n,k}$ by $p_{n,k} = \frac{|\mathcal{F}_{n,k} \cap \mathcal{P}|}{|\mathcal{F}_{n,k}|}$. Our probabilistic framework consists in taking the limit of this probability $p_{n,k}$ when the "size" (n and/or k) of the automata grows toward infinity. In [9], only a particular case was considered: k fixed, and $n \to \infty$. The following definition consider all possible enumerations of 'size' through the notion of *path*.

Definition 3.1. A path is an injective function $\rho : \mathbb{N} \to \mathbb{N}^2$. When the limit exists, we define the asymptotic density of $\mathcal{P}$ in $\mathcal{F}$ following a path ρ by

$$d_{\rho,\mathcal{F}}(\mathcal{P}) = \lim_{x \to \infty} \frac{|\mathcal{F}_{\rho(x)} \cap \mathcal{P}|}{|\mathcal{F}_{\rho(x)}|}$$

The family of possible paths is huge and two different paths do not always define different densities.

We denote $\mathbb{N}_{c_0} = \mathbb{N} \setminus \{0, 1, \ldots, c_0 - 1\}$. Since we consider asymptotics, we can restrain to paths $\rho : \mathbb{N} \to \mathbb{N}_{n_0} \times \mathbb{N}_{k_0}$ without loss of generality.

In the following, we will obtain limit densities of value 1, which justifies the use of non-cumulative density : in our case a density 1 following a given path implies a cumulative limit density 1 along this path.

3.2. Density of monotone properties among symmetric family

A property $\mathcal{P}$ is said to be *increasing* with respect to simulation if $\forall \mathcal{A} \in \mathcal{P}$, $\mathcal{A} \preccurlyeq \mathcal{B}$ implies $\mathcal{B} \in \mathcal{P}$. Decreasing properties are defined analogously. In this section we prove that monotone properties have density 0 or 1 among symmetric families introduced in section 1 following particular paths. More precisely, we are going to show that any non-trivial increasing property has density 1.

. For any local function $f : \mathcal{Q}_n^k \to \mathcal{Q}_n$, for any set $E \subseteq \mathcal{Q}_n^k$, we denote by $f|_E$ the restriction of f to E. We also denote $\mathcal{F}_{n,k}|_E = \{f|_E : f \in \mathcal{F}_{n,k}\}$. Let $\{E_i\}_{i \in I}$ be a finite family of subsets of $\mathcal{Q}_n^k$ and denote $E = \cup_{i \in I} E_i$. We say that the family $\{E_i\}_{i \in I}$ is independent for $\mathcal{F}$ if the map

$$\psi : \mathcal{F}_{n,k} \;\to\; \mathcal{F}_{n,k}|_{\mathcal{Q}_n^k \setminus E} \times \prod_{i \in I} \mathcal{F}_{n,k}|_{E_i}$$

defined by $\psi(f) = (f|_{\mathcal{Q}_n^k \setminus E}, f|_{E_1}, \ldots, f|_{E_i}, \ldots)$ is a bijection (it is always injective).

By extension, we say that a collection of subshifts $\{X_i\}_{i \in I}$ is independent if the family $\{E(X_i)\}_{i \in I}$ is independent, where $E(X_i) \subseteq \mathcal{Q}_n^k$ is the set of words of length k occurring in X_i.

Let $\mathcal{S}_{\mathcal{A}_0} = \{\mathcal{A} \in \mathbf{CA} : \mathcal{A}_0 \preccurlyeq \mathcal{A}\}$ and $\mathcal{S}_{\mathcal{A}_0,X} = \{\mathcal{A} \in \mathbf{CA} : \mathcal{A}_0 \preccurlyeq_X \mathcal{A}\}$.

Lemma 3.2. *Let $\mathcal{F} \subseteq$ **CA**, and $\mathcal{A}_0 \in \mathcal{F}_{n_0,k_0}$ a given CA. For any size (n,k) ($n \geq n_0$, $k \geq k_0$) and any collection of subshifts $\{X_i\}_{i\in I}$, we denote $\alpha_i = \frac{|\mathcal{F}_{n,k} \cap \mathcal{S}_{\mathcal{A}_0,X_i}|}{|\mathcal{F}_{n,k}|}$ for all i. If $\{X_i\}_{i\in I}$ is independent for $\mathcal{F}$, then*

$$\frac{|\mathcal{F}_{n,k} \cap \mathcal{S}_{\mathcal{A}_0}|}{|\mathcal{F}_{n,k}|}. \geq 1 - \prod_{i\in\mathcal{I}} (1 - \alpha_i)$$

Proof. We use the notations above. As the property $\mathcal{A}_0 \preccurlyeq_{X_i} \mathcal{A}$ is only determined by the restriction of $\mathcal{A}$ to $E(X_i)$, there exists $A_i \subseteq \mathcal{F}_{n,k}|_{E_i}$ such that $\psi(\mathcal{F}_{n,k} \cap \mathcal{S}_{\mathcal{A}_0,X_i}) = \mathcal{F}_{n,k}|_{\mathcal{Q}_n^k \setminus E} \times \mathcal{F}_{n,k}|_{E_1} \times \cdots \times \mathcal{F}_{n,k}|_{E_{i-1}} \times A_i \times \mathcal{F}_{n,k}|_{E_{i+1}} \cdots$. And as the family $\{E_i\}_{i\in I}$ is independent for $\mathcal{F}$, ψ is bijective and $\alpha_i = \frac{|A_i|}{|\mathcal{F}_{n,k}|_{E_i}|}$.

By definition of $\mathcal{S}_{\mathcal{A}_0}$ we have the following inclusion: $\bigcup_{i\in I}(\mathcal{F}_{n,k} \cap \mathcal{S}_{\mathcal{A}_0,X_i}) \subseteq (\mathcal{F}_{n,k} \cap \mathcal{S}_{\mathcal{A}_0})$. To conclude, it is enough to use the fact that ψ is bijective in order to express the size of these sets' complement in $\mathcal{F}_{n,k}$.

■

3.2.1. *Increasing n, fixed k.*

Proposition 3.3. *In the following, $\boldsymbol{E}$ is chosen among **CA**, **MS**, **Set**, $\boldsymbol{O}_{k'}$**Set**, $\boldsymbol{O}_{k'}$**MS**. For any $\mathcal{A}_0 \in \boldsymbol{E} \cap \boldsymbol{K}_{n_0,k_0}$, for all ϵ, there exists n_{ϵ,k_0} such that if $n \geq n_{\epsilon,k_0}$*

$$\frac{|(\boldsymbol{E} \cap \boldsymbol{K}_{n,k_0}) \bigcap \mathcal{S}_{\mathcal{A}_0}|}{|\boldsymbol{E} \cap \boldsymbol{K}_{n,k_0}|} \geq 1 - \epsilon$$

Thus, any increasing property $\mathcal{P}$ such that $\exists \mathcal{A}_0 \in \mathbf{E} \cap \mathbf{K}_{n,k} \cap \mathcal{P}$ has density 1 in family $\mathbf{E} \cap \mathbf{K}$ for paths with fixed k. The case $\mathbf{E} = \mathbf{CA}$ was already proved in [9].

Proof. Let $\{X_i\}_{i\in [\![1;\lfloor \frac{n}{n_0} \rfloor]\!]}$ be a collection of fullshifts on disjoints alphabets of size n_0. They are independent for family $\mathbf{E}\cap\mathbf{K}$, whatever the choice of $\mathbf{E}$. Because of captivity constraint, the simulation happens on X_i with probability $\alpha_{i,n,k_0} \geq c_0 = 1/n_0^{n_0^{k_0}}$. We obtain by lemma 3.2 $\frac{|(\mathbf{E}\cap\mathbf{K}_{n,k_0}) \bigcap \mathcal{S}_{\mathcal{A}_0}|}{|\mathbf{E}\cap\mathbf{K}_{n,k_0}|} \geq 1 - (1 - c_0)^{\lfloor \frac{n}{n_0} \rfloor}$.

■

3.2.2. *Increasing n, fixed k.* In the following, we use lemma 3.2, with an increasing number $l = O(k)$ of independent simulation subshifts, each providing the desired property for a constant fraction d_n of $\mathcal{F}_{n,k}$ (n is fixed). It gives $\frac{|\mathcal{F}_{n,k} \cap \mathcal{S}_{\mathcal{A}_0}|}{|\mathcal{F}_{n,k}|} \geq 1 - (1 - d_n)^l$ and we obtain a limit density $d_{k,\mathcal{F}}(\mathcal{S}_{\mathcal{A}_0}) = 1$.

Multiset CA.

Proposition 3.4. *For all $\mathcal{A}_0 \in \boldsymbol{MS}_{n_0,k_0}$, for all $\epsilon > 0$, for all $n \geq n_0 + 2$, there exists k_ϵ such that for all $k > k_\epsilon$, $\frac{|\boldsymbol{MS}_{n,k} \cap \mathcal{S}_{\mathcal{A}_0}|}{|\boldsymbol{MS}_{n,k}|} > 1 - \epsilon$.*

Proof. We consider a multiset CA $\mathcal{A}_0 \in \mathbf{MS}_{n_0,k_0}$, a size $n \geq n_0 + 2k_0 + 4$, and a given $\epsilon > 0$. In order to clarify the construction we denote the 2 biggest states of $\mathcal{Q}_n$ by 0_0 and 1_0. For any size k, we define $l = \lfloor \frac{k-k_0}{k_0-1} \rfloor$ and $o = k - l.k_0$. And for any $j \in [\![k_0 + 1; l - k_0 - 1]\!]$, M_j is the word $M_j = 0_0^{l-j} \cdot 1_0^j$.

We define the simulating subshift X_j as the set of configurations alternating a state of $\mathcal{Q}_{n_0}$ and a pattern M_j. The family $\{X_j\}_j$ is independent for multiset CA. On every such subshift, the simulation will happen if the CA maintains the structure (eventually shifted) and computes steps of $\mathcal{A}_0$. Multisets corresponding to patterns of length k occurring in X_j are:

- $V_{j,\{\!\{(x_1,1),(x_2,1),\ldots,(x_{k_0},1)\}\!\}} =$
  $\{\!\{(0_0,(k_0-1).j+o),(1_0,(k_0-1).(l-j)),(x_1,1);(x_2,1),\ldots,(x_{k_0},1)\}\!\}$
- For $0 \le s \le o$, $W^0_{j,s,k_0-1} =$
  $\{\!\{(0_0,(k_0-1).j+o+1-s),(1_0,(k_0-1).(l-j)+s),(x_1,1),(x_2,1),\ldots,(x_{k_0-1},1)\}\!\}$
- $W^1_{j,k_0-1} = \{\!\{(0_0,(k_0-1).j),(1_0,(k_0-1).(l-j)+o+1),(x_1,1),(x_2,1),\ldots,(x_{k_0-1},1)\}\!\}$
- For $0 \le s \le o-1$, $W^{1'}_{j,s,k_0} =$
  $\{\!\{(0_0,(k_0-1).j+s),(1_0,(k_0-1).(l-j)+o-s),(x_1,1),(x_2,1),\ldots,(x_{k_0},1)\}\!\}$

$\mathcal{A}_0$ is simulated on support X_j if we have the following:

- $\delta_{\mathcal{A}}(V_{j,\{\!\{(x_1,1),(x_2,1),\ldots,(x_{k_0},1)\}\!\}}) = \delta_{\mathcal{A}_0}(\{\!\{(x_1,1),(x_2,1),\ldots(x_{k_0},1)\}\!\})$
- $\delta_{\mathcal{A}}(W^0_{j,s,k_0-1}) = 0_0$ with $0 \le s \le o$
- $\delta_{\mathcal{A}}(W^1_{j,k_0-1}) = \delta_{\mathcal{A}}(W^{1'}_{j,s,k_0}) = 1_0$ with $0 \le s \le o-1$

The number of involved legal multiset transitions for a given subshift X_j is less than $(2.k_0+1).n_0^{k_0}$. Thus, the proportion of CA in $\mathbf{MS}_{n,k}$ simulating $\mathcal{A}_0$ on X_j is at least $1/n^{(2.k_0+1).n_0^{k_0}}$ which is constant with increasing k. And the number of such possible subshift is $l = O(k)$. We conclude with lemma 3.2 as explained before. ■

Totalistic CA. We manage to make the multiset construction above to become totalistic. To do it, we define the mapping φ_j by: $\forall x \in \mathcal{Q}_{n_0}$, $\varphi_j(x) = (x(n_0+1)) \cdot 0_0^{l-j} \cdot 1_0^j$, with $0_0 = 0$ and $1_0 = n_0(n_0+1)+1$. The j-th subshift is defined as the smallest subshift containing $\left(\varphi_j(\mathcal{Q}_n^k)\right)^{\mathbb{Z}}$. The transitions are distinguishable by the number of 1_0, and the number of states smaller than $n_0(n_0+1)$ in any legal neighbourhood. The probability to simulate the original CA on the j-th subshift is constant, and the simulating subshifts are independent for totalistic CA. As the number of possible simulation increases, the limit probability for any CA to simulate a given CA is increasing to 1.

Outer-multiset CA. We still consider the same *possible* simulations of any multiset CA $\mathcal{A}_0 \in \mathbf{CA}_{n_0,k_0}$ by a CA $\mathbf{O}_{k'}\mathbf{MS}_{n,k}$.

As $\mathcal{A}$ is only partially multiset, the number of transitions involved in a simulation on one given subshift has increased: we have to consider the transitions with every possible central pattern of size k'. Using a precise account, we ensure that the number of transitions involved in one given simulation is bounded by $c^{k'}$ with c only depending on n_0 and k_0. And the number of possible subshifts for the simulation to happen is the same as in the totally multiset case: it is still given by $\lfloor k/2 \rfloor - 1$. We obtain $\frac{|\mathbf{O}_{k'}\mathbf{MS}_{n,k} \cap \mathcal{S}_{\mathcal{A}_0}|}{|\mathbf{O}_{k'}\mathbf{MS}_{n,k}|} > 1 - \left(1 - \frac{1}{(n_0+2)^{c^{k'}}}\right)^l$ with $l = O(k)$. To ensure that $d_{k,\mathbf{O}_{k'}\mathbf{MS}_{n,k}}(\mathcal{S}_{\mathcal{A}_0}) = 1$ it is enough to suppose that $k' = o(log(log(k))$.

3.2.3. *More general paths.*

Multiset captive CA. We prove a slightly more general result with the family of multiset captive CA **KMS** defined by $\mathbf{KMS} = \mathbf{K} \cap \mathbf{MS}$.

Proposition 3.5. *For any path* $\rho : \mathbb{N} \to \mathbb{N}^2$ *such that* the lower limit of $x \mapsto n = \pi_1(\rho(x))$ is infinite, *and for any* $\mathcal{A}_0 \in \boldsymbol{KMS}_{n_0,k_0}$, *for all* ϵ, *there exists* s_ϵ *such that if* $x > s_\epsilon$ *then*

$$\frac{|\mathcal{S}_{\mathcal{A}_0} \cap \boldsymbol{KMS}_{\rho(x)}|}{|\boldsymbol{KMS}_{\rho(x)}|} > 1 - \epsilon$$

Proof. The collection of subshifts, and the simulation behaviour are exactly the same as in the multiset case. If $\mathcal{A}_0$ is captive, each simulating transition is also captive. The number of involved transitions is the same as in the **MS** case: $(2.k_0 + 1).n_0^{k_0}$. But using the captivity constraint, the probability for the simulation on the j-th subshift to happen is also bounded by : $1/(2.k_0 + 1).n_0^{k_0}$. We use the fact that the number of possible simulations is still $O(k)$ to conclude using lemma 3.2. ∎

Set captive CA.

Proposition 3.6. *For any path* $\rho : \mathbb{N} \to \mathbb{N}^2$ *such that* the lower limit of $x \to n = \pi_1(\rho(x))$ is infinite, *and for any* $\mathcal{A}_0 \in \boldsymbol{KSet}_{n_0,k_0}$, *for all* ϵ, *there exists* s_ϵ *such that if* $x > s_\epsilon$ *then*

$$\frac{|\mathcal{S}_{\mathcal{A}_0} \cap \boldsymbol{KSet}_{Path(x)}|}{|\boldsymbol{KSet}_{Path(x)}|} > 1 - \epsilon$$

Proof. Given $\mathcal{A}_0$, n, and k big enough, we denote the $2k_0 + 4$ first states of $\mathcal{Q}_n$ by 0_i and 1_i, $i \in [\![1; k_0 + 2]\!]$. The j-th subshift is the set of configurations alternating words $0_i^o 1_i^{l-o}$ (with $l = \lfloor \frac{k-k_0}{k_0-1} \rfloor$ and $o = k - l.k_0$) legally ordered and simulating states taken from $\Sigma_j = [\![2k_0 + 4 + j.n_0; 2k_0 + 4 + j.n_0 + n_0 - 1]\!]$. Legal set transitions for this subshift are

- $\{a_1, \ldots, a_{k_0}\} \cup \{\underline{0_i}, 1_i, \ldots, 0_{i+k-1}, 1_{i+k-1}, \underline{0_{i+k}}\} \to \delta_{\mathcal{A}_0}(\{a_1, \ldots, a_k\})$
- $\{a_1, \ldots, a_{k_0+e}\} \cup \{\underline{1_{i-1}}, 0_i, 1_i, \ldots, 0_{i+k-1}, \underline{1_{i+k-1}}, \underline{0_{i+k}}\} \to 1_{i+k/2}$ with $e \in \{0, -1\}$
- $\{a_1, \ldots, a_{k_0+e}\} \cup \{\underline{0_i}, \underline{1_i}, 0_{i+1}, 1_{i+1}, \ldots, 1_{i+k-1}, 0_{i+k}, \underline{1_{i+k}}\} \to 0_{i+k/2}$ with $e \in \{0, -1\}$

With indicies modulo $k + 2$, and $a_x \in \Sigma_j$ for all x. For all i those transitions may be identified by a set CA using the underlined state.

. So we need $n_0 + 2.(k_0 + 2)$ different states to make the simulation on this subshift and the number of involved transitions is equal to $3.(k_0 + 2)$. Thus, because of captivity, the proportion p of CA in which one given simulation happens is constant when k, or n is increasing. And the family of the $\lfloor \frac{n-2(k_0+2)}{n_0} \rfloor$ possible simulation subshifts is independent. With lemma 3.2, we obtain the inequality $\frac{|\mathcal{S}_{\mathcal{A}_0} \cap \mathbf{KSet}_{n,k}|}{|\mathbf{KSet}_{n,k}|} > 1-(1-p)^{\lfloor \frac{n-2(k_0+2)}{n_0} \rfloor}$. We conclude the proof using the hypothesis on the path, $\underline{lim}_{x\to\infty} n = \underline{lim}_{x\to\infty} \pi_1(x) = \infty$. ∎

4. Encodings

In the following we prove that there exists universal cellular automata in most of the families defined above. This is an important step considering the fact that some well known locally defined family, such as LR-permutative CA, do not contain any universal CA (because intrinsic universality implies non-surjectivity, see [5]). In fact, for every given family $\mathcal{F}$, we introduce an encoding map $\varphi_{\mathcal{F}} : \mathbf{CA} \to \mathcal{F}$ such that for any $\mathcal{A}$, its corresponding encoded version $\varphi_{\mathcal{F}}(\mathcal{A})$ verifies $\mathcal{A} \preccurlyeq \varphi_{\mathcal{F}}(\mathcal{A})$. The existence of a universal CA in $\mathcal{F}$ follows by application of the encoding to any universal CA. Moreover, in some cases, we obtain a stronger result: the encoded CA is universal if and only if the original CA is universal.

Set CA. Given a CA $\mathcal{A} \in \mathbf{CA}_{n,k}$ of state set $\mathcal{Q}_n$, we construct $\Psi(\mathcal{A}) \in \mathbf{Set}$ with state set $Q = \mathcal{Q}_n \times \{0, \dots, k+1\} \cup \{\#\}$ of size $n.(k+2)+1$.

A configuration $c \in Q^{\mathbb{Z}}$ is said *legal* if $c(z) \neq \#$ for all z and if the projection of c on the second component of states (which is well-defined) is periodic of period $1 \cdot 2 \cdots (k+2)$. Thus, for any legal configuration c and any position z, the set of states of cells which are neighbours of z is of the form:

$$E_i(a_1, \dots, a_k) = \{(a_1, i), (a_2, i+1 \ mod\ k+2), ..., (a_k, i+k-1 \ mod\ k+2)\}$$

for some $i \in \{1, \dots, k+2\}$ (with $a_j \in \mathcal{Q}_n$ for all j). $\Psi(\mathcal{A})$ is defined by the local rule f as follows:

$$f(x_1, \dots, x_k) = \begin{cases} \big(\delta_{\mathcal{A}}(a_1, \dots, a_k), i + \lfloor k/2 \rfloor \ mod\ k+2\big) & \text{if } \{x_1, \dots, x_k\} = E_i(a_1, \dots, a_k), \\ \# & \text{else.} \end{cases}$$

By construction, we have $\Psi(\mathcal{A}) \in \mathbf{Set}$. Moreover the encoding preserves universality. As a direct corollary, we get the undecidability of universality in family **Set** (universality was proven undecidable in the general case in [6]).

Theorem 4.1. *The encoding Ψ : **CA** $\to$ **Set** satisfies the following:*
(1) *$\mathcal{A} \preccurlyeq \Psi(\mathcal{A})$ for all $\mathcal{A}$,*
(2) *$\mathcal{A}$ is universal if and only if $\Psi(\mathcal{A})$ is universal.*

Captive set CA. We denote by **KSet** the intersection $\mathbf{K} \cap \mathbf{Set}$. The previous construction does not generally produce captive CA (even if the original CA is captive). We now describe a new encoding which produces only CA belonging to **KSet**. It could have been used to prove the existence of universal set CA, but we have no proof that it satisfies the second assertion of theorem 4.1 (hence the usefulness of previous construction).

The new mapping $\varphi : \mathbf{CA} \to \mathbf{KSet}$ is an adaptation of Ψ. We keep the idea of states being a cartesian product of the original alphabet $\mathcal{Q}_n$ and a family of labels which is in this case $\{0, ..., 2k-2\}$. But, in order to have every transition satisfying the captive constraint, we introduce 'libraries' of states placed regularly in legal configurations: between two computing cells, we place the i-th library for some i, denoted by $\mathbf{L}_i$, which contains the n states $\{(x,i)\}_{x \in \mathcal{Q}_n}$. For technical reasons, it also contains special states $(\#, i)$ and $(\#', i)$, and it is ordered as follows: $\mathbf{L}_i = (\#, i) \cdot (1, i) \cdot (2, i) \cdots (n, i) \cdot (\#', i)$. Thus, $\varphi(\mathcal{A})$ has state set $Q = \{0, \dots, 2k-2\} \times (\mathcal{Q}_n \cup \{\#, \#'\})$.

The simulation of $\mathcal{A}$ by $\varphi(\mathcal{A})$ takes place on 'legal' configurations defined by an alternation of an isolated state of label i, and a library of type $k+i$, precisely:

$$\cdots\ (q_1, i)\ \cdot\ \mathbf{L}_{k+i \bmod 2k-1}\ \cdot\ (q_2, i+1 \bmod 2k-1)\ \cdot\ \mathbf{L}_{k+i+1 \bmod 2k-1} \cdots$$

Those legal configurations will be maintained in one-to-one correspondence with configurations of $\mathcal{A}$, successive isolated states between libraries corresponding to successive states from $\mathcal{A}$. However, this time, the simulation of 1 step of $\mathcal{A}$ will use 2 steps of $\varphi(\mathcal{A})$ and only even time steps of $\varphi(\mathcal{A})$ (including time 0) will produce legal configurations. For odd time steps, we introduce 'intermediate' configurations defined by an alternation of an isolated state of label i, and a library of type $r+i$, precisely:

$$\cdots\ (q_1, i)\ \cdot\ \mathbf{L}_{r+i \bmod 2k-1}\ \cdot\ (q_2, i+1 \bmod 2k-1)\ \cdot\ \mathbf{L}_{r+i+1 \bmod 2k-1} \cdots$$

where $r = \lfloor k/2 \rfloor$ is the radius of $\mathcal{A}$.

To describe the local rule of $\varphi(\mathcal{A})$, we introduce the following sets:

- $V_i(a_1, \ldots, a_k) = \{(a_1, i), (a_2, i+1 \bmod 2k-1) \ldots (a_k, i+k-1 \bmod 2k-1)\}$;
- L_i is the set of states present in the word $\mathbf{L}_i$;
- $B_{i,x} = \{(\#, i), (1, i), ..., (b-1, i)\}$ is the set of states in the prefix of of $\mathbf{L}_i$ of length x;
- $E_{i,x} = \{(e, i), ..., (n, i), (\#', i)\}$ is the set of states in the suffix of $\mathbf{L}_i$ of length $n-x+1$.

$\varphi(\mathcal{A})$ has arity $k' = k + (k-1) \cdot (n+2)$ and, on legal configurations, the set of states seen in a neighbourhood has one of the following types:

T1: $V_i(a_1, \ldots, a_k) \cup L_{i+k \ mod\ 2k-1} \cup \ldots \cup L_{i+2k-2 \ mod\ 2k-1}$;
T2: $V_i(a_1, \ldots, a_k) \cup E_{i+k-1 \ mod\ 2k-1, x} \cup L_{i+k \ mod\ 2k-1} \cup \ldots$
$\ldots \cup L_{i+2k-3 \ mod\ 2k-1} \cup B_{i+2k-2 \ mod\ 2k-1, x}$.

On intermediate configurations, the set of states seen in a neighbourhood has one of the following types:

T3: $V_i(a_1, \ldots, a_k) \cup L_{i-r \ mod\ 2k-1} \cup \ldots \cup L_{i-r+k-2 \ mod\ 2k-1}$;
T4: $V_i(a_1, \ldots, a_k) \cup E_{i-r-1 \ mod\ 2k-2, x} \cup L_{i-r \ mod\ 2k-1} \cup \ldots$
$\ldots \cup L_{i-r+k-3 \ mod\ 2k-1} \cup B_{i-r+k-2 \ mod\ 2k-1, x}$.

An important point is that the 4 types are disjoint: it is obvious that each of T1 and T3 is disjoint from each of T2 and T4, and the overall disjointness follows from the fact that sets of type T3 and T4 have less elements than T1 and T2 since set L_i are disjoint but

$$V_i(a_1, \ldots, a_k) \cap L_j \neq \emptyset \iff i \leq j \leq i+k-1$$

Using notations above, the behaviour of $\varphi(\mathcal{A})$ is defined by 4 kinds of transitions according to the kind of neighbourhood seen:

T1: $\to (\delta_{\mathcal{A}}(a_1, ...a_k), i+2k-2 \bmod 2k-1)$
T2: $\to (x, i+k-1 \bmod 2k-1)$
T3: $\to (a_{1+r}, i+r \bmod 2k-1)$
T4: $\to (x, i+r)$

The crucial point for transition of type T3 to be well-defined is that a_{1+r} can be unambiguously determined given that the libraries present have labels ranging from $i-r-1$ to $i-r+k-2 = i+r-1$ whereas a_{1+r} is associated to label $i+r$ in $V_i(a_1, \ldots, a_k)$ (everything is taken modulo $2k-1$).

Intuitively, type T1 corresponds to simulation of transitions of $\mathcal{A}$ and the three other types are devoted to the modification of label of isolated states or the displacement of libraries according to the following scheme:

- At even steps, transitions of type T1 apply the local rule $\delta_{\mathcal{A}}$, but the result receive a label j such that $\mathbf{L}_j$ is present in the neighbourhood to satisfy captivity constraint; meanwhile, transitions of type T2 just shift the libraries.
- At odd steps, the difference of labels between libraries and isolated states is wrong; to come back to a legal configuration, transitions of type T3 leave isolated states unchanged while transitions of type T4 shift the libraries.

To completely define $\varphi(\mathcal{A})$, we fix some ordering on Q and specify that, when the set E of neighbours doesn't correspond to any of the 4 types above, the local rule of $\varphi(\mathcal{A})$ simply chooses the greatest state in E. With that definition, $\varphi(\mathcal{A})$ always belong to **KSet**, because it depends only on the set of states in the neighbourhood, and because each transition produces a state already present in the neighbourhood (either the neighbourhood contains L_i for the right value of i, or the local rule simply chooses the greatest element).

Theorem 4.2. *For any $\mathcal{A}$, we have $\mathcal{A} \preccurlyeq \varphi(\mathcal{A})$. Therefore families **MS**, **KMS**, **Set** and **KSet** contain universal CA.*

The construction above corresponds to the strongest symmetry constraint (captive set CA), put aside totalistic CA. The existence of (intrinsically) universal totalistic CA is proven in [5]. The case of outer-totalistic CA follows by inclusion.

5. Universality Everywhere

Gathering the density results of section 3 and the existential results for universality in section 4, we obtain an asymptotic density 1 for universality in the following classes.

Family $\mathcal{F}$	**Condition on the path ρ**	**Comments**
Captive CA	$\exists k_0$ s.t. $\rho(x) = (x, k_0)$	Already in [9]
Multiset CA	$\exists n_0$ s.t. $\rho(x) = (n_0, x)$	
k'-outer-multiset	$\exists n_0$ s.t. $\rho(x) = (n_0, x)$	$k' = o(\log(\log k))$
Totalistic CA	$\exists n_0$ s.t. $\rho(x) = (n_0, x)$	
k'-outer-totalistic	$\exists n_0$ s.t. $\rho(x) = (n_0, x)$	$k' = o(\log(\log k))$
Set captive CA	$\underline{\lim}_{x\to\infty}\ \pi_1(\rho(x)) = +\infty$	
Multiset captive CA	$\underline{\lim}_{x\to\infty}\ \pi_1(\rho(x)) = +\infty$	

6. Open Problems and Future Work

As summarised in the previous section, our work establishes that universality has asymptotic density 1 along path ρ in several families defined by local symmetries, provided ρ verifies some hypothesis depending on the family considered.

Notably, we leave open the question of the density of universality in the following cases:

- increasing state set for families **MS**, **Set**, **Tot** (and outer-versions),
- increasing neighbourhood for family **K**.

We have no result (and no intuition) concerning the case of the whole set of CA either. A possible progress on that topic could be to reduce the density problem of a family $\mathcal{F}_1$ to the density problem of a family $\mathcal{F}_2$, *i.e.* to show that the densities (if they exist) in the two families are equal up to non-trivial multiplicative constants.

Another perspective, especially for multiset CA (or sub-families **Set** and **Tot**), is to extend our result to higher dimensions or even to more general lattice of cells. Indeed, the symmetry involved here implies isotropy which is an often required property in modelling.

Finally, it remains to study typical dynamics obtained in each family from random initial configuration. Experiments suggest that self-organisation in those families is far more frequent than in CA in general.

References

[1] Bastien Chopard and Michel Droz. *Cellular automata modeling of physical systems*. Collection Aléa-Saclay: Monographs and Texts in Statistical Physics. Cambridge University Press, Cambridge, 1998.

[2] J. Kari. Reversibility and Surjectivity Problems of Cellular Automata. *Journal of Computer and System Sciences*, 48(1):149–182, 1994.

[3] P. Kůrka. Languages, equicontinuity and attractors in cellular automata. *Ergodic Theory and Dynamical Systems*, 17:417–433, 1997.

[4] J. Mazoyer and I. Rapaport. Inducing an Order on Cellular Automata by a Grouping Operation. In *Symposium on Theoretical Aspects of Computer Science*. Lecture Notes in Computer Science, 1998.

[5] N. Ollinger. *Automates Cellulaires : structures*. PhD thesis, École Normale Supérieure de Lyon, dcembre 2002.

[6] N. Ollinger. The intrinsic universality problem of one-dimensional cellular automata. In *Symposium on Theoretical Aspects of Computer Science*, pages 632–641. Lecture Notes in Computer Science, 2003.

[7] Nicolas Ollinger. Universalities in cellular automata: a (short) survey. In B. Durand, editor, *Symposium on Cellular Automata Journées Automates Cellulaires (JAC'08)*, pages 102–118. MCCME Publishing House, Moscow, 2008.

[8] G. Theyssier. Captive cellular automata. In *Mathematical Foundations of Computer Science*, pages 427–438. Lecture Notes in Computer Science, 2004.

[9] G. Theyssier. How common can be universality for cellular automata? In *Annual Symposium on Theoretical Aspects of Computer Science*, 2005.

[10] Wolfram. Statistical mechanics of cellular automata. *Review of Modern Physics*, 55, 1983.

[11] S. Wolfram. Universality and complexity in cellular automata. *Physica D*, 10:1–35, 1984.